

LOCAL SYMMETRY IN RANDOM GRAPHS

JEFFERSON ELBERT SIMÕES

DANIEL R. FIGUEIREDO

VALMIR C. BARBOSA

*Systems Engineering and Computer Science Program, COPPE
Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*

ABSTRACT. Quite often real-world networks can be thought of as being symmetric, in the abstract sense that vertices can be found to have similar or equivalent structural roles. However, traditional measures of symmetry in graphs are based on their automorphism groups, which do not account for the similarity of local structures. We introduce the concept of *local symmetry*, which reflects the structural equivalence of the vertices' egonets. We study the emergence of asymmetry in the Erdős-Rényi random graph model and identify regimes of both asymptotic local symmetry and asymptotic local asymmetry. We find that local symmetry persists at least to an average degree of $n^{1/3}$ and local asymmetry emerges at an average degree not greater than $n^{1/2}$, which are regimes of much larger average degree than for traditional, global asymmetry.

1. INTRODUCTION

Graphs have become some of the most versatile mathematical objects, capable of representing a wide range of real-world structures such as the Internet, neural networks and scientific collaboration networks, to name a few. A natural question to pose in many of these scenarios is: what is the meaning of symmetry in the context of graphs? Several definitions have been proposed for this term, each more adequate to a certain context or application. Usually, such definitions are based upon transformations over graphs that preserve certain properties of their structures, the most traditional one being the existence of non-trivial automorphisms. Such automorphisms allow us to classify the vertices of a graph according to their “role” in its structure.

The concept of symmetry is closely tied to that of *structural identity* [21], which is the identification of vertices based on features of the network structure and their relationships to other vertices. Structural identity can be applied in several contexts, such as network privacy [1]. For instance, data on social networks including personal information is usually anonymized by the removal of labels, in an attempt to preserve the privacy of their members (vertices). If such network is symmetric, vertices with equal structural roles cannot be distinguished without some kind of

E-mail addresses: {elbert,daniel}@land.ufrj.br, valmir@cos.ufrj.br.

This work has been partially funded by CAPES, CNPq, and FAPERJ BBP grants.

side information, which makes them more likely to resist attempts at deanonymization [13].

However, in several applied contexts, the role of a vertex in the network structure, and as a consequence its structural identity, is intuitively related to its vicinity in the network, rather than the whole network. For instance, in neural networks (in which vertices correspond to neurons, and edges to the synapses between them), different areas of the brain become responsible for specific functions, such as memory or motor coordination [22]. In another example, in networks such as the web, nodes can be assigned roles such as *hub* or *authority*, depending on how they are connected to their own vicinities [17].

Our contributions are as follows. We propose a definition of *local symmetry*, based on the structural similarity of neighborhoods around each vertex. Our definition naturally induces a hierarchy of symmetries, which progressively use more information for classifying vertices, ultimately culminating in the traditional, automorphism-based symmetry, which we call *global symmetry* in the context of this work. Furthermore, we study the emergence of our base form of symmetry in the context of the Erdős-Rényi random graph. We find that, relative to global symmetry, asymptotic local symmetry occurs for graphs with much larger average degrees, in particular for degrees growing as fast as $n^{1/3}$ (Theorem 3.1), while asymptotic local asymmetry eventually emerges, at most at degrees slightly larger than $n^{1/2}$ (Theorem 5.1).

2. LOCAL SYMMETRY

The abstract concept of symmetry is traditionally embodied in the context of graphs by the notions of *isomorphism* and *automorphism*. Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are said to be *isomorphic* if there is a function $f : V_1 \rightarrow V_2$, called an *isomorphism* between G_1 and G_2 , that is bijective and precisely maps the edges of G_1 into edges of G_2 . This is a known equivalence relation on the set of all graphs, and it can be thought of as identifying graphs with their edge structure, ignoring the nature of their vertices or the interpretation of their edges. Indeed, this view is implicit in traditional graph theory expressions such as “up to isomorphism”.

An isomorphism between a graph $G = (V, E)$ and G itself is called an *automorphism* of G . Every graph has a so-called *trivial* automorphism given by the identity function \mathbb{I}_V over V , and a graph having no other automorphisms is said to be *asymmetric* [9]. The set of automorphisms of a graph possesses a group structure when coupled with the operation of composition, which induces equivalence classes in its vertex set. Again, there is a natural interpretation to this fact: vertices can be grouped according to their placement in the graph structure, such that vertices in the same class are “structurally indistinguishable”, at least without additional information about the identities of (potentially all) the remaining vertices. This interpretation leads us to the following definitions:

Definition 2.1. *Given a graph $G = (V, E)$, two vertices $v_1, v_2 \in V$ are globally symmetric if there is an automorphism f of G such that $f(v_1) = v_2$.*

Definition 2.2. *Let $G = (V, E)$ be a graph. Then G is said to be globally symmetric if there are $u, v \in V$ distinct and globally symmetric.*

Remark 2.3. *G is globally symmetric if and only if it has at least one non-trivial automorphism.*

Throughout this paper, we will use the term “globally asymmetric” both for pairs of vertices and for graphs that are not globally symmetric. Alternative definitions, specifically in the literature of probability theory, have employed the simpler term

symmetric graph for graphs satisfying Remark 2.3 [9], but this conflicts with at least two known graph-theoretic definitions for the term “symmetric graph” [2, p.104] [14]. Thus, introducing a new term in this definition may help avoid ambiguities.

One of the greatest difficulties in studying symmetry in graphs lies in quantifying what it means for a graph to be “almost symmetric”. The usefulness of such quantification comes from the fact that, in many real-world networks, one can identify intuitively equivalent vertices that the definition of global symmetry falls short in capturing. For instance, consider the network in Figure 1. Intuitively, the vertices u and v can be seen as small “hubs”, who share almost equally the role of connecting the remaining, “peripheral” vertices, which are also intuitively equivalent to each other. However, since u and v have different degrees, no automorphism is able to map them onto one another, or map a peripheral neighbor of u onto a peripheral neighbor of v . Thus, the equivalence classes of vertices found on this graph do not capture this intuitive notion of symmetry.

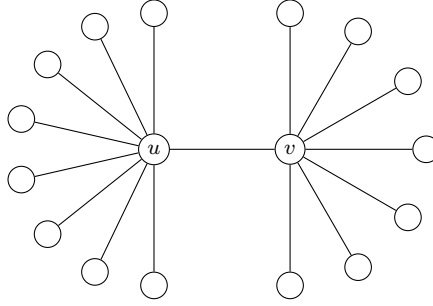


FIGURE 1. Example of “almost symmetric” graph. Vertices u and v connect their neighbors to the remainder of the graph.

While several measures have been proposed as a more flexible notion of symmetry [9, 6], no proposal seems to have achieved wide acceptance, because of both the computational complexity of calculating such measures and the absence of knowledge about their relationship to real-world networks and applications.

A second difficulty, which we address in this work and that has barely been analyzed in the literature, lies in identifying vertices that (again intuitively) have equivalent local structures but are globally distinguishable. The graph in Figure 2 illustrates this idea well. In this figure, we highlight two isomorphic induced subgraphs. However, since both these subgraphs are part of a larger graph, the existence of an automorphism that reflects such symmetry depends on the edge pattern of the remainder of the graph, and in this particular example the desired automorphism does not exist. In other words, even though these subgraphs are intuitively symmetric, since they are connected differently to the graph, they are not globally symmetric.

Note that, regardless of the structure of the remainder of the graph, and despite the fact that the intuitive equivalence between the two local structures continues to exist, the two aforementioned difficulties can happen simultaneously. For instance, in the example of Figure 2, by removing the single edge identified by a dashed line and located at least at distance 5 from the highlighted subgraphs, the graph acquires an automorphism that maps these subgraphs precisely. Hence, the existence of this automorphism — a global mapping satisfying local restrictions — may be sensitive to changes that seem unpretentious and unrelated to the local structures of interest.

To proceed more precisely, we must first delineate which regions we are interested in analyzing. Given a graph $G = (V, E)$ and a vertex set $S \subseteq V$, we define $\mathcal{N}_G(S)$,

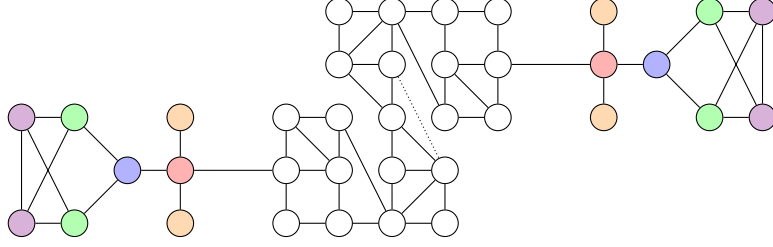


FIGURE 2. Example of equivalence between local structures in a graph.

the *open neighborhood* of S in G :

$$\mathcal{N}_G(S) = \{v \in V : d(v, S) \leq 1\},$$

and the *closed neighborhood* of S in G :

$$\mathcal{N}_G[S] = G[\mathcal{N}_G(S)],$$

where, for every set $A \subseteq V$ of vertices, $d(v, A)$ is the smallest distance between v and some vertex of A , and $G[A]$ is the subgraph induced by the vertices of A . Naturally, $\mathcal{N}_G(\{v\})$ (or, by simplicity, $\mathcal{N}_G(v)$) is the set that comprises v and its neighbors, and $\mathcal{N}_G[\{v\}]$ (or $\mathcal{N}_G[v]$) is the subgraph induced by v and its neighbors. We will omit the index G when the graph at hand is clear.

This definition of neighborhood is traditional in the field of graph theory¹, and we can extend it to include not only vertices at distance 1, but at distance k . We achieve this goal easily by defining the *open k -neighborhood* of S in G :

$$\mathcal{N}_G^k(S) = \{v \in V : d(v, S) \leq k\}.$$

Note that $\mathcal{N}^0(S) = S$, $\mathcal{N}^1(S) = \mathcal{N}(S)$ and, recursively, $\mathcal{N}^k(S) = \mathcal{N}(\mathcal{N}^{k-1}(S))$. We define the *closed k -neighborhood* of S in G similarly:

$$\mathcal{N}_G^k[S] = G[\mathcal{N}_G^k(S)].$$

We illustrate these definitions in Figure 3. Note that $\mathcal{N}^k(S) \subseteq \mathcal{N}^{k+1}(S)$ and that $\mathcal{N}^k[S]$ is an induced subgraph of $\mathcal{N}^{k+1}[S]$. We also note that, in contexts such as mathematical sociology, the closed neighborhood of a vertex is known as an *egonet* [7, 4].

For the purposes of this work, we use closed k -neighborhoods around single vertices as a proxy for “locality”. This motivates the following definition:

Definition 2.4. *Given a graph $G = (V, E)$, two vertices $v_1, v_2 \in V$ are k -locally symmetric if there is an isomorphism f between $\mathcal{N}^k[v_1]$ and $\mathcal{N}^k[v_2]$ such that $f(v_1) = v_2$.*

Therefore, two vertices v_1 and v_2 are k -locally symmetric if the k -th order local structures in which v_1 and v_2 are located are equivalent, with v_1 and v_2 equivalently located in these structures. If $k = 1$, we will briefly say that the vertices are *locally symmetric*.

One interesting feature of k -local symmetry is that it naturally leads to the construction of a symmetry hierarchy, which includes global symmetry as the most restrictive one, as evinced by the following result.

Proposition 2.5. *Let v_1 and v_2 be vertices of $G = (V, E)$, and let $k \in \mathbb{N}$. Then the following statements hold:*

¹Actually, the convention in graph theory is to define $\mathcal{N}_G(S) = \{v \in V : d(v, S) = 1\}$. The reason we choose this alternate definition should be clear in the remainder of this text.

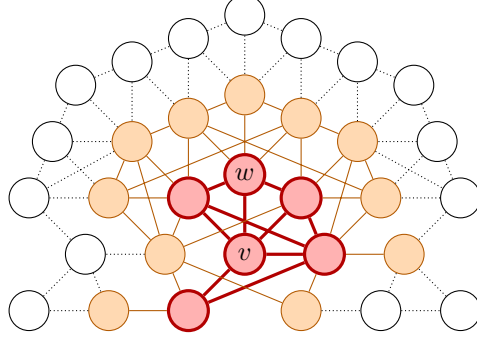


FIGURE 3. Examples of neighborhoods of a vertex in a graph. The subgraph highlighted in red is $\mathcal{N}^1[v] = \mathcal{N}[v]$, and including the orange highlights we obtain $\mathcal{N}^2[v]$. The vertex sets of these subgraphs are, respectively, $\mathcal{N}^1(v) = \mathcal{N}(v)$ and $\mathcal{N}^2(v)$. For $k \geq 3$, $\mathcal{N}^k[v] = G$.

- (1) If v_1 and v_2 are $(k+1)$ -locally symmetric, then v_1 and v_2 are k -locally symmetric;
- (2) If $k \geq \text{diam}(G)$, then v_1 and v_2 are k -locally symmetric if and only if v_1 and v_2 are globally symmetric.

Proof. If v_1 and v_2 are $(k+1)$ -locally symmetric, then there is an isomorphism f between $\mathcal{N}^{k+1}[v_1]$ and $\mathcal{N}^{k+1}[v_2]$ with $f(v_1) = v_2$. Since isomorphisms preserve distances, for any $d \in \mathbb{N}$, two vertices u and v of $\mathcal{N}^{k+1}[v_1]$ are at distance d if and only if $f(u)$ and $f(v)$ are at distance d in $\mathcal{N}^{k+1}[v_2]$. Therefore, for every vertex w in $\mathcal{N}^{k+1}[v_2]$, it is true that $w \in \mathcal{N}^k[v_1] \iff f(w) \in \mathcal{N}^k[v_2]$. This means that $f|_{\mathcal{N}^k[v_1]}$ is a bijection between $\mathcal{N}^k(v_1)$ and $\mathcal{N}^k(v_2)$. Note that, since f also preserves edges, so does $f|_{\mathcal{N}^k[v_1]}$, thus $f|_{\mathcal{N}^k[v_1]}$ is an isomorphism between $\mathcal{N}^k[v_1]$ and $\mathcal{N}^k[v_2]$. Since $f|_{\mathcal{N}^k[v_1]}(v_1) = f(v_1) = v_2$, it follows that v_1 and v_2 are, by definition, k -local symmetric.

This proves the first statement. The second statement follows from the fact that, if $k \geq \text{diam}(G)$, then any two vertices of G are at distance k or smaller from each other. Therefore, $\mathcal{N}^k(v) = V$ and $\mathcal{N}^k[v] = G$, and the definitions of k -local symmetry and global symmetry are equivalent. \square

Finally, we define local symmetry in graphs:

Definition 2.6. Let $G = (V, E)$ be a graph. Then G is k -locally symmetric if there are $u, v \in V$ distinct and k -locally symmetric.

Note that this definition for k -local symmetry is analogous to Definition 2.2 for global symmetry. Therefore, our definitions of k -local symmetry — for two vertices and for a single graph — are mutually consistent in the same fashion as those of global symmetry. This definition also implies a symmetry hierarchy similar to Proposition 2.5:

Proposition 2.7. Let $G = (V, E)$ be a graph, and let $k \in \mathbb{N}$. Then the following statements hold:

- (1) If G is $(k+1)$ -locally symmetric, then G is k -locally symmetric;
- (2) If $k \geq \text{diam}(G)$, then G is k -locally symmetric if and only if G is globally symmetric.

Given our interest in understanding symmetry in random network models, it is natural that we start our analysis with the $G(n, p)$ model. In particular, we would

like to determine if there is any fundamental difference between the emergence of local symmetry and global symmetry in this model. It is known [9] [3, p.230] that, for the similar $G(n, m)$ model of random graphs, an asymmetric graph is obtained a.a.s. if and only if $2m/\binom{n}{2} \geq \log n + \omega(1)$ and $n - 1 - 2m/\binom{n}{2} \geq \log n + \omega(1)$ — that is, when the model exhibits average degree at least slightly larger than $\log n$ and at most slightly smaller than $n - 1 - \log n$. For $G(n, p)$, an analogous but slightly weaker result is known: the $G(n, p)$ random graph is asymmetric a.a.s. if $p \in [\log n/n, 1 - \log n/n]$ [15], and it is globally symmetric if $p \ll \log n/n$ (due to the existence of isolated vertices a.a.s. [12]) or $1 - p \ll \log n/n$ (due to the existence of universal vertices a.a.s.).

3. SYMMETRY REGIMES

Our first result is the identification of a local symmetry regime for $G(n, p)$:

Theorem 3.1. *A $G(n, p)$ random graph, with $p = o(n^{-2/3})$, is locally symmetric a.a.s.*

Let us proceed with some terminology before the proof of this statement. Recall the definition of closed 1-neighborhood of a vertex. We call this vertex the *center* of this subgraph, with all edges between the center and other vertices said to be *core edges*, and all remaining edges termed *peripheral edges*.

The idea behind the proof is that peripheral edges always close triangles, but in this regime, $G(n, p)$ does not have too many triangles, so the closed neighborhood of many vertices are simple stars. Two such vertices will have isomorphic closed neighborhoods simply by having the same degree, which will happen for some pair, since the degrees in $G(n, p)$ concentrate heavily around their mean.

To execute the proof, we will need two auxiliary results:

Lemma 3.2. *Let $G = (V, E)$ be a $G(n, p)$ random graph. If $p = \omega(\log n/n)$, then for any fixed $\delta \in (0, 1)$, the degree of all vertices of G are within the range $(n - 1)p(1 \pm \delta)$, a.a.s.*

Proof. Let d_v be the degree of vertex v . We know that $d_v \stackrel{d}{\sim} \text{Bin}(n - 1, p)$. By the Chernoff bound, for any $\delta \in (0, 1)$:

$$\mathbb{P}[d_v \notin (n - 1)p(1 \pm \delta)] \leq 2e^{-(n-1)p\delta^2/3}.$$

The union bound implies:

$$\mathbb{P}[\exists v : d_v \notin (n - 1)p(1 \pm \delta)] \leq 2ne^{-(n-1)p\delta^2/3}$$

and, since $np = \omega(\log n)$ by hypothesis, the right-hand side is $2ne^{-\omega(\log n)} = 2no(1/n) = o(1)$. \square

Lemma 3.3. *Let $G = (V, E)$ be a $G(n, p)$ random graph, and let T be the number of triangles in G . Then, $\mathbb{E}[T] = \binom{n}{3}p^3$ and, if $p = \omega(1/n)$, $\mathbb{P}[|T - \mathbb{E}[T]| < c \cdot \mathbb{E}[T]] \rightarrow 1$ for any fixed $c > 0$.*

Proof. We denote by $\binom{V}{3}$ the set of unordered triples of vertices, and for each triple $t = (i, j, k)$, we define the event $\{t \text{ is } \Delta\} = \{(i, j), (i, k), (j, k) \in E\}$. Then:

$$T = \sum_{t \in \binom{V}{3}} \mathbb{I}_{\{t \text{ is } \Delta\}}.$$

It easily follows from linearity of expectation and independence of edges that $\mathbb{E}[T] = \sum_{t \in \binom{V}{3}} \mathbb{P}[t \text{ is } \Delta] = \binom{n}{3}p^3$.

We also need to estimate the variance of T , which we denote by $\mathbb{V}[T]$. For such, we need an expression for its second moment:

$$\begin{aligned}
\mathbb{E}[T^2] &= \mathbb{E} \left[\left(\sum_{t \in \binom{V}{3}} \mathbb{I}_{\{t \text{ is } \Delta\}} \right) \left(\sum_{t' \in \binom{V}{3}} \mathbb{I}_{\{t' \text{ is } \Delta\}} \right) \right] \\
&= \sum_{t, t' \in \binom{V}{3}} \mathbb{E}[\mathbb{I}_{\{t \text{ is } \Delta\}} \mathbb{I}_{\{t' \text{ is } \Delta\}}] \\
&= \sum_{t, t' \in \binom{V}{3}} \mathbb{P}[t \text{ is } \Delta, t' \text{ is } \Delta]
\end{aligned}$$

This summation can be broken into four pieces, based on the relationship between the two triples of vertices, t and t' :

No common vertices: All edges of t are independent of all edges of t' , so $\mathbb{P}[t \text{ is } \Delta, t' \text{ is } \Delta] = p^6$;

One common vertex: Again, edges of t are independent of edges of t' , and $\mathbb{P}[t \text{ is } \Delta, t' \text{ is } \Delta] = p^6$;

Two common vertices: t and t' share the edge between common vertices, comprising a total of 5 edges — thus, $\mathbb{P}[t \text{ is } \Delta, t' \text{ is } \Delta] = p^5$;

Three common vertices: In this case, $t = t'$ and $\mathbb{P}[t \text{ is } \Delta, t' \text{ is } \Delta] = p^3$.

We must also count how many triples fit into each of these cases:

No common vertices: $\binom{n}{3} \binom{n-3}{3}$ triples;

One common vertex: $\binom{n}{3} \cdot 3 \binom{n-3}{2}$ triples;

Two common vertices: $\binom{n}{3} \cdot 3(n-3)$ triples;

Three common vertices: $\binom{n}{3}$ triples.

We can now calculate the second moment of T :

$$\begin{aligned}
\mathbb{E}[T^2] &= \sum_{t, t' \in \binom{V}{3}} \mathbb{P}[t \text{ is } \Delta, t' \text{ is } \Delta] \\
&= \binom{n}{3} p^3 + 3 \binom{n}{3} (n-3) p^5 + 3 \binom{n}{3} \binom{n-3}{2} p^6 + \binom{n}{3} \binom{n-3}{3} p^6 \\
&= \binom{n}{3} p^3 \left[1 + 3np^2 - 9p^2 + \frac{3}{2}n^2p^3 - \frac{3}{2}7np^3 + \frac{3}{2}12p^3 \right. \\
&\quad \left. + \frac{1}{6}n^3p^3 - \frac{1}{6}12n^2p^3 + \frac{1}{6}47np^3 - \frac{1}{6}60p^3 \right].
\end{aligned}$$

We can also obtain an expression for $\mathbb{E}[T]^2$:

$$\begin{aligned}
\mathbb{E}[T]^2 &= \left(\binom{n}{3} p^3 \right)^2 \\
&= \binom{n}{3} p^3 \left[\frac{1}{6}n^3p^3 - \frac{1}{6}3n^2p^3 + \frac{1}{6}2np^3 \right].
\end{aligned}$$

Combining these expressions yields an expression for $\mathbb{V}[T]$:

$$\begin{aligned}
\mathbb{V}[T] &= \mathbb{E}[T^2] - \mathbb{E}[T]^2 \\
&= \binom{n}{3} p^3 (1 + 3np^2 - 9p^2 - 3np^3 + 8p^3)
\end{aligned}$$

Note that, if $np = \omega(1)$, then $\mathbb{V}[T]/\mathbb{E}[T]^2 \rightarrow 0$ when $n \rightarrow \infty$. Chebyshev's inequality states that, for any $k > 0$,

$$\mathbb{P}[|T - \mathbb{E}[T]| \geq k\sqrt{\mathbb{V}[T]}] \leq \frac{1}{k^2}.$$

Then, for any $c > 0$, we can take $k = c \mathbb{E}[T]/\sqrt{\mathbb{V}[T]}$ and obtain

$$\mathbb{P}[|T - \mathbb{E}[T]| \geq c \mathbb{E}[T]] \leq \frac{\mathbb{V}[T]}{c^2 \mathbb{E}[T]^2},$$

which vanishes for fixed c . \square

We can now proceed to the proof of Theorem 3.1.

Proof of Theorem 3.1. We will first prove the result under the additional assumption that $p = \omega(\log n/n)$, which will be removed by the end of the proof.

Let $G = (V, E)$ be a $G(n, p)$ random graph, and let T be the number of triangles in G . Fix $c > 0$ arbitrary and $\delta \in (0, 1)$, and define the following sequences of events on G :

$$A_n = \{|T - \mathbb{E}[T]| < c \mathbb{E}[T]\},$$

$$B_n = \{\text{all degrees} \leq (n-1)p(1+\delta)\}.$$

Lemma 3.2 and Lemma 3.3 ensure that $\mathbb{P}(A_n \cap B_n) \rightarrow 1$ as $n \rightarrow \infty$. We will prove that this intersection event is contained in the event $\{G(n, p) \text{ is locally symmetric}\}$.

In A_n , there are at most $(1+c)\binom{n}{3}p^3$ triangles in G . Each edge in a triangle appears as a peripheral edge in the closed neighborhood of its opposite vertex in this triangle. Therefore, summing over all vertices' closed neighborhoods, there are, at most, $3(1+c)\binom{n}{3}p^3 = o(n)$ peripheral edges. This implies that, at least, $n - 3(1+c)\binom{n}{3}p^3$ vertices have no peripheral edges in their closed neighborhoods. Let C be the set of such vertices.

In B_n , every vertex has degrees in the range $[0, (n-1)p(1 \pm \delta)]$. Let D be the set of integers satisfying this property.

Now, for $p = o(n^{-2/3})$, we have:

$$\begin{aligned} |C| &= n - \theta((np)^3) \\ &= n - o(n), \\ |D| &\leq (n-1)p(1+\delta) + 1 \\ &= o(\sqrt[3]{n}). \end{aligned}$$

This implies that $|C| > |D|$ for sufficiently large n . By the pigeonhole principle, in $A_n \cap B_n$, there must be at least two vertices in C with the same degree. These vertices' closed neighborhoods are stars of the same size, therefore they must be isomorphic. This implies our result.

It still remains for us to lift the assumption that $p = \omega(\log n/n)$. The main issue to resolve is that we cannot bound directly the size of D , since Lemma 3.2 does not apply. Instead, we must try and sidestep the issue to make the argument work.

Pick some $p' \geq p$, such that $\omega(\log n/n) \leq p' \leq o(n^{-2/3})$, and consider the sequence of events:

$$B'_n = \{\text{all degrees} \leq (n-1)p'(1+\delta)\}.$$

Note that our probability measure \mathbb{P} is associated with $G(n, p)$, not with $G(n, p')$. However, since B'_n happens a.a.s. under $G(n, p')$ (by Lemma 3.2), and the events

B'_n represent a decreasing property², it must also happen a.a.s. under $G(n, p)$, so $\mathbb{P}(B'_n) \rightarrow 1$.

Replacing B_n by B'_n in our previous argument, the conclusion again follows. \square

4. DEGREE FUNCTION

For the identification of asymmetry regimes, we need several additional tools. The core concept is that of the degree sequence of a graph, which we present in a slightly different form:

Definition 4.1. *For any graph $G = (V, E)$, the degree function of G is the function $\phi_G : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ such that $\phi_G(k) = |\{v \in V : d_G(v) = k\}|$ for all $k \in \mathbb{N}_0$.*

The degree function of G simply returns, for an input k , the number of vertices with degree k in G . This makes it equivalent to the degree sequence of G , whenever the listing order of the degrees is irrelevant.

To identify asymmetry regimes, we must identify conditions under which no two vertices in a $G(n, p)$ random graph are symmetric a.a.s. The asymmetry of vertices is defined as the lack of an isomorphism between their closed neighborhoods, which relates to degree functions via the following remark:

Remark 4.2. *For G and G' isomorphic graphs, $\phi_G \equiv \phi_{G'}$.*

To enable a more fine-grained look into these closed neighborhoods, we need one additional definition:

Definition 4.3. *Let $G = (V, E)$ and $G' = (V', E')$ be two graphs. The degree sequence edit distance between G and G' (denoted by $\Delta(G, G')$) is given by:*

$$\Delta(G, G') = \sum_k |\phi_G(k) - \phi_{G'}(k)|.$$

Remark 4.4. *Let μ be the counting measure on \mathbb{N} . Then:*

$$\Delta(G, G') = \int |\phi_G - \phi_{G'}| d\mu = \|\phi_G - \phi_{G'}\|_1.$$

Thus Δ is a semimetric over the space of all graphs, with $\Delta(G, G') = 0$ iff $\phi_G \equiv \phi_{G'}$.

Remark 4.5. *For G and G' isomorphic graphs, $\Delta(G, G') = 0$.*

To better understand the meaning of this edit distance, consider any partial mapping between the vertex sets of two graphs. We can measure the *degree mismatch count* of this mapping, which is a simple count of vertices, from both graphs, that are either mapped to vertices with different degrees or left unmapped.

Definition 4.6. *Let $G = (V, E)$, $G' = (V', E')$ be two graphs, and let $f : S \rightarrow S'$ be a bijective function from $S \subseteq V$ to $S' \subseteq V'$. The degree mismatch count of f (denoted by δ_f) is given by*

$$\begin{aligned} \delta_f = & |\{v \in S : d_G(v) \neq d_{G'}(f(v))\}| + |V \setminus S| \\ & + |\{v' \in S' : d_{G'}(v') \neq d_G(f^{-1}(v'))\}| + |V' \setminus S'|. \end{aligned}$$

The degree sequence edit distance between two graphs is, then, the smallest possible degree mismatch count between their vertex sets:

²A *decreasing property* is a property preserved under removal of edges (such as “ $G(n, p)$ is not connected”). A standard coupling argument shows that, for all n , the probability of such properties is a decreasing function of p .

Theorem 4.7. *Let $G = (V, E)$ and $G' = (V', E')$ be two graphs. Then:*

$$\Delta(G, G') = \min_{\substack{f : S \rightarrow S' \\ f \text{ bijective}, S \subseteq V, S' \subseteq V'}} \delta_f.$$

Proof. The statement follows from inequalities on both directions. We begin by showing $\Delta(G, G') \geq \min_f \delta_f$. It is enough to construct a function g such that $\Delta(G, G') = \delta_g$, and we perform this construction by “slices”, one for each possible vertex degree.

For each $k \in \mathbb{N}_0$, let $v_1^k, \dots, v_{\phi_G(k)}^k$ and $w_1^k, \dots, w_{\phi_{G'}(k)}^k$ be enumerations of degree- k vertices in G and G' , respectively. Write $m_k = \min(\phi_G(k), \phi_{G'}(k))$, $V_k = \{v_1^k, \dots, v_{m_k}^k\}$ and $V'_k = \{w_1^k, \dots, w_{m_k}^k\}$, and construct function $g_k : V_k \rightarrow V'_k$ mapping v_j^k to w_j^k , for $j = 1, \dots, m_k$. Note that g_k leaves $|\phi_G(k) - \phi_{G'}(k)|$ degree- k nodes unmapped, all from V (if $\phi_G(k) \geq \phi_{G'}(k)$) or from V' (if $\phi_G(k) \leq \phi_{G'}(k)$), and which, by construction, cannot be mapped by any other function $g_{k'}$.

Now, construct function $g : \cup_k V_k \rightarrow \cup_k V'_k$ as $g = \cup_k g_k$. By double counting the number of nodes left unmapped by g (from both V and V'), we see that this number is equal to $|V \setminus (\cup_k V_k)| + |V' \setminus (\cup_k V'_k)|$ by definition, and to $\sum_k |\phi_G(k) - \phi_{G'}(k)|$ by construction. Furthermore, our construction also ensures that for every node mapped by g with degree k in G , its image has degree k in G' , and vice-versa for nodes in G' . Therefore, it holds that $|\{v \in \cup_k V_k : d_G(v) \neq d_{G'}(g(v))\}| = |\{v \in \cup_k V'_k : d_{G'}(v) \neq d_G(g^{-1}(v))\}| = 0$, and:

$$\begin{aligned} \Delta(G, G') &= \sum_k |\phi_G(k) - \phi_{G'}(k)| \\ &= \sum_k |\phi_G(k) - \phi_{G'}(k)| + 0 \\ &= |V \setminus (\cup_k V_k)| + |V' \setminus (\cup_k V'_k)| \\ &\quad + |\{v \in \cup_k V_k : d_G(v) \neq d_{G'}(g(v))\}| \\ &\quad + |\{v' \in \cup_k V'_k : d_{G'}(v') \neq d_G(g^{-1}(v'))\}| \\ &= \delta_g. \end{aligned}$$

Now, it only remains to show $\Delta(G, G') \leq \min_f \delta_f$. We will show that $\Delta(G, G') \leq \delta_h$ for any partial mapping h , and we will again proceed by “slices” in our argument. Let $h : S \rightarrow S'$ be an arbitrary bijection with $S \subseteq V$ and $S' \subseteq V'$.

Consider all vertices with degree k in both G and G' . If $\phi_G(k) \geq \phi_{G'}(k)$, then at most $\phi_{G'}(k)$ k -degree vertices in G can be mapped by h into k -degree nodes in G' . This implies that at least $\phi_G(k) - \phi_{G'}(k)$ k -degree vertices in G must either be left unmapped by g (thus belonging to $V \setminus S$) or be mapped to vertices in G' with degree different than k (and therefore belonging to $\{v \in \cup_k V_k : d_G(v) \neq d_{G'}(h(v))\}$). Analogously, if $\phi_G(k) \leq \phi_{G'}(k)$, at least $\phi_{G'}(k) - \phi_G(k)$ k -degree vertices in G' must either be left unmapped by g (this time, belonging in $V' \setminus S'$) or be mapped to vertices in G with degree different than k .

In both cases, there is a contribution of $|\phi_G(k) - \phi_{G'}(k)|$ to δ_h coming exclusively from k -degree nodes in G and G' . Putting together contributions from all “slices” ensures that $\delta_h \geq \sum_k |\phi_G(k) - \phi_{G'}(k)| = \Delta(G, G')$. Since h was arbitrary, $\min_h \delta_h \geq \Delta(G, G')$, as desired. \square

This property can be used to relate the degree functions of a graph and its subgraphs. For any set of vertices $S \subseteq V$, denote by $G[S]$ the subgraph of G induced by S , and by $C(S)$ the set of edges with one endpoint in S and another in $V \setminus S$.

Corollary 4.8. *For any graph $G = (V, E)$ and any $S \subset V$,*

$$\Delta(G, G[S]) \leq |V \setminus S| + |C(S)|.$$

Proof. By virtue of Theorem 4.7, it is enough to show that $\delta_g = |V \setminus S| + |C(S)|$ for some partial mapping g between G and $G[S]$.

Take $g : S \rightarrow S$ to be the identity function on S . Then g is a partial mapping between the vertex sets of G and $G[S]$, and its degree mismatch count is given by:

$$\delta_g = |\{v \in G[S] : d_{G[S]}(v) \neq d_G(v)\}| + |V \setminus S|.$$

Now, notice that, for any vertex $v \in S$, $d_{G[S]}(v) \neq d_G(v)$ if and only if v is adjacent to some vertex outside S . Since there are $|C(S)|$ edges between S and $V \setminus S$, there can be at most $|C(S)|$ such vertices, and the result follows. \square

5. ASYMMETRY REGIMES

We can now proceed to the identification of a local asymmetry regime for the $G(n, p)$ random graph model.

Theorem 5.1. *A $G(n, p)$ random graph with $\omega(n^{-1/2+\delta_1}) \leq p \leq o(n^{-3/7-\delta_2})$ for constant $\delta_1, \delta_2 > 0$ is locally asymmetric a.a.s.*

Once again, we present some intermediate results before proceeding to the proof of this result.

Lemma 5.2. *Let X be a $\text{Bin}(n, p)$ random variable with $p < 1/2$. If $\varepsilon > 0$ and $np\varepsilon^2 \geq 3$, then*

$$\mathbb{P}[|X - np| \geq \varepsilon np] \geq 2 \exp\{-9np\varepsilon^2\}.$$

Proof. See Lemma 5.2 of [16]. \square

Lemma 5.3. *Let \vec{X} be a multinomial random vector $\text{Mult}(n, p_1, \dots, p_k)$ with k fixed, and let $0 < \beta < 1$ be also fixed. If $\Omega(n^{-\beta}) \leq p_1, \dots, p_{k-1} \leq o(1)$, then*

$$\max_{\vec{x}} \mathbb{P}[\vec{X} = \vec{x}] = O(n^{-(k-1)(1-\beta)/2}).$$

Proof. Let $x^* = (x_1^*, \dots, x_k^*)$ be the mode of $\text{Mult}(n, p_1, \dots, p_k)$. It is known [11] that $x_i^* = I(np_i)$, where $I(a)$ is either $\lfloor a \rfloor$ or $\lceil a - 1 \rceil$. This implies that $x_i^* \leq np_i \leq x_i^* + 1$. Also, since $\beta < 1$, it holds that $np_i \rightarrow \infty$ for all i , which implies that, for large enough n , $x_i^* \geq 1$ for all i .

Using these inequalities, Stirling's approximation, and known bounds for the exponential function, we have, for large enough n :

$$\begin{aligned}
\max_{\vec{x}} \mathbb{P}[\vec{X} = \vec{x}] &= \mathbb{P}[\vec{X} = x^*] \\
&= \frac{n!}{\prod_{i=1}^n x_i^*!} \prod_{i=1}^n p_i^{x_i^*} \\
&\leq \frac{en^n e^{-n} \sqrt{n}}{(\sqrt{2\pi})^k \prod_{i=1}^k (x_i^*)^{x_i^*} \sqrt{x_i^*} e^{-x_i^*}} \prod_{i=1}^k p_i^{x_i^*} \\
&\leq \frac{e\sqrt{n}}{(\sqrt{2\pi})^k} \prod_{i=1}^k \left(\frac{np_i}{x_i^*} \right)^{x_i^*} \frac{1}{\sqrt{x_i^* \cdots x_{k-1}^*}} \frac{1}{\sqrt{np_k}} \\
&\leq \frac{e}{(\sqrt{2\pi})^k} \prod_{i=1}^k \left(1 + \frac{1}{x_i^*} \right) \frac{1}{\sqrt{np_1 \cdots np_{k-1}}} \frac{1}{\sqrt{p_k}} \\
&\leq \frac{e}{(\sqrt{2\pi})^k} \cdot e^k \cdot \frac{1}{\sqrt{np_1 \cdots np_{k-1}}} \frac{1}{\sqrt{p_k}}.
\end{aligned}$$

Now, the following inequalities hold for large enough n . First, by hypothesis, for every $i \leq k-1$, we have $p_i \geq c_i n^{-\beta}$ for some constant c_i . Second, $p_k \geq c_k$ for some constant c_k , since the hypotheses imply that $p_k \rightarrow 1$. Thus:

$$\begin{aligned}
\max_{\vec{x}} \mathbb{P}[\vec{X} = \vec{x}] &\leq \frac{e}{(\sqrt{2\pi})^k} \cdot e^k \cdot \frac{1}{\sqrt{np_1 \cdots np_{k-1}}} \frac{1}{\sqrt{p_k}} \\
&\leq \frac{e^{k+1}}{(\sqrt{2\pi})^k} \frac{1}{\sqrt{c_1 \cdots c_{k-1} \cdot c_k}} \frac{1}{(\sqrt{n^{1-\beta}})^{k-1}} \\
&= Kn^{-(k-1)(1-\beta)/2}
\end{aligned}$$

for some constant K . This concludes the proof. \square

Theorem 5.4 (Theorem 3.2 in [8]). *Let $\vec{D}_n^{(1)}$ and $\vec{D}_n^{(2)}$ be the degree sequences of two independent $G(n, p)$ random graphs, with probability distribution $\mathbb{P}_{\mathcal{D}_n}$.*

Furthermore, let \mathcal{F}_n be the σ -algebra generated by $\vec{D}_n^{(1)}$ and $\vec{D}_n^{(2)}$, and let $\mathbb{P}_{\mathcal{B}_n}$ be a probability measure under which $\vec{D}_n^{(1)}$ and $\vec{D}_n^{(2)}$ are random vectors with n independent coordinates, each having distribution $\text{Bin}(n-1, p)$.

Then, for any sequence of events A_n measurable under $\mathcal{F}_n^{\otimes 2}$ and any fixed $a > 0$, if $\mathbb{P}_{\mathcal{B}_n}^{\otimes 2}(A_n) = o(n^{-a})$, then $\mathbb{P}_{\mathcal{D}_n}^{\otimes 2}(A_n) = o(n^{-a})$.

In a nutshell, Theorem 5.4 allows us to consider the degree sequence of two $G(n, p)$ random graphs as sequences of independent random variables, without interfering with power-law decays in the probability of events on this model.

Theorem 5.5. *Let G_1 and G_2 be independent $G(n, p)$ random graphs, with p satisfying $\omega(\log n/n) \leq p \leq o(n^{-1/2})$. Then, for any $\varepsilon > 0$ and any $a > 0$,*

$$\Delta(G_1, G_2) \geq n^{1/2-\varepsilon}$$

with probability $1 - o(n^{-a})$.

Proof. This proof will proceed as follows. Instead of looking at the whole degree sequences of G_1 and G_2 , we will group several ranges of degrees into “buckets”, according to their distances to the expected degree of $G(n, p)$. This will allow us to bound the probability that each vertex belongs to each bucket, using Chernoff-like bounds. Considering the degrees as independent random variables characterizes the bucketed degree sequences as multinomial random vectors. Furthermore, the

buckets themselves are carefully chosen so that the distribution of these vectors is not too concentrated, i.e., the probability of their modes is large enough. This means these two vectors will most likely be far apart from each other in L_1 -norm, which is the desired result.

We begin by noting that $\Delta(G_1, G_2) = \int |\phi_{G_1} - \phi_{G_2}| \, d\mu$ is a function of the degree sequences of G_1 and G_2 . Let ϕ'_{G_1}, ϕ'_{G_2} be the degree functions obtained by approximating these degree sequences by sequences of n independent $\text{Bin}(n-1, p)$ random variables. By virtue of Theorem 5.4, it is enough to show that

$$\int |\phi'_{G_1} - \phi'_{G_2}| \, d\mu \geq n^{1/2-\varepsilon}$$

with probability $1 - o(n^{-a})$.

Fix real positive numbers $\alpha > \sqrt{27}$ and $\beta < \varepsilon$, and choose a natural number $b > 2a/(\varepsilon - \beta)$. Let the real positive intervals S_1, \dots, S_b be defined as

$$S_i = [\alpha^{-(i+1)} \sqrt{(n-1)p} \cdot f(n, p), \alpha^{-i} \sqrt{(n-1)p} \cdot f(n, p)],$$

where $f(n, p) = (\min\{\log n, (n-1)p\})^{1/4}$. For ease the notation, let the set $S_{b+1} = \mathbb{R}_+ \setminus \cup_{i=1}^b S_i$ contain the remainder of the positive real line.

We now use these sets to group the vertices in both graphs into “buckets”, according to their degrees, with set S_i indicating the set of allowed degrees according to their distances to the average degree $(n-1)p$. More formally, define the sets of integers B_1, \dots, B_{b+1} as

$$B_i = \{k \in \mathbb{N}_0 : |k - (n-1)p| \in S_i\}.$$

Note that $\cup_i B_i = \mathbb{N}_0$ and $B_i \cap B_j = \emptyset$ whenever $i \neq j$. Now, write $T_i^{(1)} = \int_{B_i} \phi'_{G_1} \, d\mu$. $T_i^{(1)}$ counts the number of vertices of G_1 with degrees in B_i . Similarly, write $T_i^{(2)} = \int_{B_i} \phi'_{G_2} \, d\mu$. Then it holds that

$$\begin{aligned} \int |\phi'_{G_1} - \phi'_{G_2}| \, d\mu &= \sum_{i=1}^{b+1} \int_{B_i} |\phi'_{G_1} - \phi'_{G_2}| \, d\mu \\ &\geq \sum_{i=1}^{b+1} \left| \int_{B_i} \phi'_{G_1} \, d\mu - \int_{B_i} \phi'_{G_2} \, d\mu \right| \\ &= \sum_{i=1}^{b+1} |T_i^{(1)} - T_i^{(2)}| = \|\vec{T}^{(1)} - \vec{T}^{(2)}\|_1, \end{aligned}$$

where $\vec{T}^{(j)} = (T_1^{(j)}, \dots, T_{b+1}^{(j)})$.

This means that

$$\begin{aligned} \mathbb{P} \left[\int |\phi'_{G_1} - \phi'_{G_2}| \, d\mu \geq n^{1/2-\varepsilon} \right] &\leq \mathbb{P} \left[\|\vec{T}^{(1)} - \vec{T}^{(2)}\|_1 \geq n^{1/2-\varepsilon} \right] \\ &= \sum_{\vec{t}} \mathbb{P} \left[\vec{T}^{(1)} = \vec{t}, \|\vec{T}^{(2)} - \vec{t}\|_1 \geq n^{1/2-\varepsilon} \right] \\ &= \sum_{\vec{t}} \mathbb{P} \left[\vec{T}^{(1)} = \vec{t} \right] \mathbb{P} \left[\|\vec{T}^{(2)} - \vec{t}\|_1 \geq n^{1/2-\varepsilon} \right] \\ &\leq \sum_{\vec{t}} \mathbb{P} \left[\vec{T}^{(1)} = \vec{t} \right] \max_{\vec{t}} \mathbb{P} \left[\|\vec{T}^{(2)} - \vec{t}\|_1 \geq n^{1/2-\varepsilon} \right] \\ &= \max_{\vec{t}} \mathbb{P} \left[\|\vec{T}^{(2)} - \vec{t}\|_1 \geq n^{1/2-\varepsilon} \right]. \end{aligned}$$

Note that, for any \vec{t} , the event $\{\|\vec{T}^{(2)} - \vec{t}\|_1 \geq n^{1/2-\varepsilon}\}$ has at most $(2n^{1/2-\varepsilon} + 1)^b$ elements: each of the first b coordinates of $\vec{T}^{(2)}$ must be at distance at most

$n^{1/2-\varepsilon}$ from the corresponding coordinate of \vec{t} (for a maximum of $2n^{1/2-\varepsilon} + 1$ valid options), and the last one is uniquely determined from the previous choices, since the coordinates must sum up to n . This means that:

$$\begin{aligned} \mathbb{P} \left[\int |\phi'_{G_1} - \phi'_{G_2}| \, d\mu \geq n^{1/2-\varepsilon} \right] &\leq \max_{\vec{t}} \mathbb{P} \left[\|\vec{T}^{(2)} - \vec{t}\|_1 \geq n^{1/2-\varepsilon} \right] \\ &\leq (2n^{1/2-\varepsilon} + 1)^b \max_{\vec{t}} \mathbb{P} \left[\vec{T}^{(2)} = \vec{t} \right]. \end{aligned}$$

Now, the degree of every vertex in G_1 or G_2 must belong to some B_i , and these degrees are deemed to be i.i.d. random variables, by our initial argument regarding ϕ'_{G_1} and ϕ'_{G_2} . This implies that $\vec{T}^{(1)}, \vec{T}^{(2)}$ are multinomial random vectors $\text{Mult}(n, p_1, \dots, p_{b+1})$, where $p_i = \mathbb{P}[\text{Bin}(n-1, p) \in B_i]$ is the probability that the degree of a vertex belongs to B_i .

At this moment, to apply Lemma 5.3, we would like to bound from both sides the value of p_i (for $i \leq b$). For an upper bound, an application of the Chernoff bound suffices:

$$\begin{aligned} p_i &= \mathbb{P}[\text{Bin}(n-1, p) \in B_i] \\ &\leq \mathbb{P}[|\text{Bin}(n-1, p) - (n-1)p| \geq \alpha^{-(i-1)} \sqrt{(n-1)p} \cdot f(n, p)] \\ &\leq 2 \exp\{-f(n, p)^2 \alpha^{-2(i+1)} / 3\}, \end{aligned}$$

which is $o(1)$, since $f(n, p) \rightarrow \infty$.

For a lower bound, we use the Chernoff bound and an application of Lemma 5.2, noting that $\alpha^{-2(i+1)} f(n, p)^2 \geq 3$ for large enough n :

$$\begin{aligned} p_i &= \mathbb{P}[\text{Bin}(n-1, p) \in B_i] \\ &= \mathbb{P}[|\text{Bin}(n-1, p) - (n-1)p| \geq \alpha^{-(i-1)} \sqrt{(n-1)p} \cdot f(n, p)] \\ &\quad - \mathbb{P}[|\text{Bin}(n-1, p) - (n-1)p| \geq \alpha^{-i} \sqrt{(n-1)p} \cdot f(n, p)] \\ &\geq 2 \exp\{-9\alpha^{-2(i+1)} f(n, p)^2\} - 2 \exp\{-\alpha^{-2i} f(n, p)^2 / 3\} \\ &= 2 \exp\{-9\alpha^{-2(i+1)} f(n, p)^2\} (1 - \exp\{-\alpha^{-2(i+1)} f(n, p)^2\}^\gamma), \end{aligned}$$

where, in the last passage, we let $\gamma = \alpha^2/3 - 9 > 0$. Note that $\gamma > 0$ and $f(n, p) \rightarrow \infty$ imply that the term inside the parentheses tends to 1. The remaining exponential satisfies

$$\begin{aligned} n^\beta \exp\{-9\alpha^{-2(i+1)} f(n, p)^2\} &\geq \exp\{\beta \log n - 9\alpha^{-2(i+1)} \sqrt{\log n}\} \\ &= \exp\{\beta \log n (1 - 9\alpha^{-2(i+1)} (\log^{-1/2} n))\} \\ &= \omega(1), \end{aligned}$$

thus the expression on the right-hand side is $\omega(n^{-\beta})$.

This means that the conditions of Lemma 5.3 are satisfied for random vectors $\vec{T}^{(1)}$ and $\vec{T}^{(2)}$, with $k = b + 1$. Therefore, $\max_{\vec{t}} \mathbb{P}[\vec{T}^{(2)} = \vec{t}] = O(n^{-b(1-\beta)/2})$, and, by the choices of b and β ,

$$\begin{aligned} \mathbb{P} \left[\int |\phi'_{G_1} - \phi'_{G_2}| \, d\mu \geq n^{1/2-\varepsilon} \right] &\leq (2n^{1/2-\varepsilon} + 1)^b \max_{\vec{t}} \mathbb{P} [\vec{T}^{(2)} = \vec{t}] \\ &= \theta(n^{b(1/2-\varepsilon)}) O(n^{-b(1-\beta)/2}) \\ &= O(n^{-b(\varepsilon-\beta)/2}) \\ &= o(n^{-a}). \end{aligned}$$

□

With this, we can proceed to the proof of Theorem 5.1.

Proof of Theorem 5.1. Let $G = (V, E)$ be a $G(n, p)$ random graph. By the union bound, it is enough to prove that any two distinct vertices v_1, v_2 are locally symmetric with probability $o(n^{-2})$.

Let $v_1, v_2 \in V$ be arbitrary distinct vertices, and denote by X_1 the set of neighbors of v_1 that are not neighbors of v_2 , by X_2 the set of neighbors of v_2 that are not neighbors of v_1 , and by Y the set of common neighbors of v_1 and v_2 . Additionally, denote by C_1 the number of edges between X_1 and Y , and by C_2 the number of edges between X_2 and Y . Our goal is to show that, with probability $o(n^{-2})$, $\mathcal{N}[v_1] = G[X_1 \cup Y]$ and $\mathcal{N}[v_2] = G[X_2 \cup Y]$ are not isomorphic. Note that these two graphs are not independent, since they share $G[Y]$ as a subgraph. Our goal is to show that, even if $G[Y]$ is removed from these two subgraphs, their degree sequences have large enough edit distance, by Theorem 5.5, that the effect of reinserting $G[Y]$, which is bounded by Corollary 4.8, is not enough to make these degree sequences equal, which implies $\mathcal{N}[v_1]$ and $\mathcal{N}[v_2]$ cannot possibly be isomorphic.

Before proceeding to this, we will need a few concentration bounds that will help us carry out our proof. First, note that $v_2 \in X_1$ and $v_1 \in X_2$ if and only if u and v are neighbors, and that all other vertices belong to X_1 , X_2 and Y independently from each other, with probabilities $p(1-p)$, $p(1-p)$ and p^2 , respectively. Thus, $|X_1|, |X_2| \leq_d \text{Bin}(n-2, p) + 1$ and $|Y| \stackrel{d}{\sim} \text{Bin}(n-2, p^2)$. By a similar reasoning, given X_1 , X_2 , and Y , it holds that $C_1 \stackrel{d}{\sim} \text{Bin}(|X_1||Y|, p)$ and $C_2 \stackrel{d}{\sim} \text{Bin}(|X_2||Y|, p)$.

Now, set $0 < \varepsilon_1 < 1$ constant, $\varepsilon_2 = \log n$, and $\varepsilon_3 > 0$ constant. Define the following events:

$$\begin{aligned} A_1 &= \{(n-2)p(1-p)(1-\varepsilon_1) < |X_1|, |X_2| < (n-2)p(1-p)(1+\varepsilon_1)\}, \\ A_2 &= \{|Y| < (n-2)p^2(1+\varepsilon_2)\}, \\ A_3 &= \{C_1, C_2 < sp(1+\varepsilon_3)\}, \end{aligned}$$

where $s = (n-2)^2 p^3 (1-p)(1+\varepsilon_1)(1+\varepsilon_2)$.

We will show that $\mathbb{P}(A_1, A_2, A_3) \geq 1 - o(n^{-2})$. First, by the Chernoff bound,

$$\begin{aligned} \mathbb{P}(\overline{A_1}) &\leq 2 \exp\{-(n-2)p(1-p)\varepsilon_1 \min(1, \varepsilon_1)/3\} \\ &\quad + 2 \exp\{-(n-2)p(1-p)\varepsilon_1 \min(1, \varepsilon_1)/2\}, \\ \mathbb{P}(\overline{A_2}) &\leq \exp\{-(n-2)p^2\varepsilon_2 \min(1, \varepsilon_2)/3\}. \end{aligned}$$

Furthermore, in the event $A_1 \cap A_2$, it holds that $|X_1||Y|, |X_2||Y| \leq s$, which implies $C_1, C_2 \leq_d \text{Bin}(s, p)$ and, again by the Chernoff bound:

$$\mathbb{P}(A_1, A_2, \overline{A_3}) \leq 2 \exp\{-sp\varepsilon_3 \min(1, \varepsilon_3)/3\}.$$

Note that the upper bounds for $\mathbb{P}(\overline{A_1})$, $\mathbb{P}(\overline{A_2})$ and $\mathbb{P}(A_1, A_2, \overline{A_3})$ are all $o(n^{-2})$, since $(n-2)p \geq \omega(\log n)$ and $(n-2)p^2\varepsilon_2 \geq \omega(1)$ by hypothesis, and $sp \geq \omega(\log n)$ as a consequence. Thus, $\mathbb{P}(A_1, A_2, A_3) \geq 1 - o(n^{-2})$.

We can now resume the main thread in our proof and show that, conditional on $A_1 \cap A_2 \cap A_3$, v_1 and v_2 are locally asymmetric with probability $1 - o(n^{-2})$. Recall that v_1 and v_2 are locally symmetric only if $\Delta(\mathcal{N}[v_1], \mathcal{N}[v_2]) = 0$. We can assume that v_1 and v_2 have the same degree — i.e. $|\mathcal{N}(v_1)| = |\mathcal{N}(v_2)|$, otherwise $\mathcal{N}[v_1]$ and $\mathcal{N}[v_2]$ have vertex sets of different sizes, which implies v_1 and v_2 are locally asymmetric. Note that this implies that $|X_1| = |X_2|$: if v_1 and v_2 are adjacent, then $\mathcal{N}(v_1) = X_1 \uplus Y \uplus \{v_2\}$ and $\mathcal{N}(v_2) = X_2 \uplus Y \uplus \{v_1\}$; if they are not adjacent, then $\mathcal{N}(v_1) = X_1 \uplus Y$ and $\mathcal{N}(v_2) = X_2 \uplus Y$.

Let $N_1 = (V_1, E_1) = G[\mathcal{N}(v_1) \setminus \{v_1\}]$ and $N_2 = (V_2, E_2) = G[\mathcal{N}(v_2) \setminus \{v_2\}]$, i.e., $\mathcal{N}[v_1]$ and $\mathcal{N}[v_2]$ with their centers removed. Since v_1 and v_2 are universal vertices of their respective neighborhoods, their removal keeps the edit distance between them unchanged, that is,

$$\Delta(\mathcal{N}[v_1], \mathcal{N}[v_2]) = \Delta(N_1, N_2).$$

Since Δ is a semimetric, it further holds that

$$\Delta(N_1, N_2) \geq \Delta(G[X_1], G[X_2]) - \Delta(G[X_1], N_1) - \Delta(G[X_2], N_2).$$

To lower bound the first term on the right-hand side, we note that the existence of edges between vertices in X_1 happens with probability p independently of the particular vertex pair, thus $G[X_1]$ is a $G(|X_1|, p)$ random graph. Since $p = \omega(n^{-1/2+\delta_1})$, by hypothesis, and $|X_1| = \theta(np)$ in $A_1 \cap A_2 \cap A_3$, it holds, for n large enough and for constants $c, c' > 0$, that

$$\begin{aligned} \frac{p}{\log |X_1|/|X_1|} &\geq \frac{p \cdot cnp}{\log c'np} \\ &\geq \frac{\omega(n^{2\delta_1})}{\log c'n} \\ &\geq \omega(1). \end{aligned}$$

It also holds that $|X_1|p^2 = p \cdot \theta(np^3) = o(1)$. Together, these inequalities imply $\omega(\log |X_1|/|X_1|) \leq p \leq o(|X_1|^{-1/2})$, and therefore $G[X_1]$ satisfies the hypotheses of Theorem 5.5. An analogous argument implies that $G[X_2]$ also satisfies these hypotheses.

Furthermore, note that $G[X_1]$ and $G[X_2]$ are independent, since their vertex sets are disjoint. This allows us to apply the results of Theorem 5.5, picking fixed $a = \frac{14}{1-14\delta_2}$ and $\varepsilon < \frac{49\delta_2}{8-14\delta_2}$:

$$\begin{aligned} \mathbb{P}[\Delta(G[X_1], G[X_2]) \geq |X_1|^{1/2-\varepsilon} |A_1 \cap A_2 \cap A_3|] &= 1 - o((|X_1|p)^{-a}) \\ &= 1 - o((np^2)^{-a}) \\ &\geq 1 - o(n^{-a(1+2(-3/7-\delta_2))}) \\ &= 1 - o(n^{-a(1/7-2\delta_2)}) \\ &= 1 - o(n^{-2}). \end{aligned}$$

Note that these choices of a and ε are always possible and valid, since the hypotheses imply $\delta_2 < 1/14$, which makes $\frac{49\delta_2}{8-14\delta_2} \geq 7\delta_2 > 0$ and $\frac{14}{4-7\delta_2} > 0$. The two bounds on $|X_1|$ from the definition of event A_1 imply that, with probability $1 - o(n^{-2})$,

$$\Delta(G[X_1], G[X_2]) \geq \Omega((np)^{1/2-\varepsilon}).$$

To upper bound the remaining two terms, we begin by applying Corollary 4.8:

$$\Delta(G[X_1], N_1) \leq |V_1 \setminus X_1| + |C(X_1)|.$$

Now, all vertices of Y are in $V_1 \setminus X_1$, and so is v_2 if it is a neighbor of v_1 in G . All other vertices of G either are in X_1 or are not in V_1 , thus $|V_1 \setminus X_1| \leq |Y| + 1$. Furthermore, $|C(X_1)|$ counts all edges from X_1 to $V_1 \setminus X_1$, regardless of whether v_2 belongs to $V_1 \setminus X_1$, since no edge connects X_1 to v_2 by construction. Thus, $|C(X_1)|$ counts edges from X_1 to Y , which implies $|C(X_1)| = C_1$. It follows that

$$\begin{aligned} \Delta(G[X_1], N_1) &\leq |Y| + 1 + C_1 \\ &\leq (n-2)p^2(1+\varepsilon_2) + (n-2)^2p^4(1-p)(1+\varepsilon_1)(1+\varepsilon_2)(1+\varepsilon_3) \\ &= \theta(n^2p^4 \log n). \end{aligned}$$

By an analogous argument, the same inequality holds for $\Delta(G[X_2], N_2)$. Now, note that

$$\begin{aligned} \frac{\Delta(G[X_1], N_1) + \Delta(G[X_2], N_2)}{\Delta(G[X_1], G[X_2])} &\leq \frac{\theta(n^2 p^4 \log n)}{\Omega((np)^{1/2-\varepsilon})} \\ &= o(n^{3/2+\varepsilon} p^{7/2+\varepsilon} \log n) \\ &\leq o(n^{3/2+\varepsilon} n^{(-3/7-\delta_2)(7/2+\varepsilon)} \log n) \\ &= o(n^{(4-7\delta_2)\varepsilon/7-7\delta_2/2} \log n) \end{aligned}$$

The exponent of n in this expression satisfies

$$\begin{aligned} \frac{(4-7\delta_2)\varepsilon}{7} - \frac{7\delta_2}{2} &< \frac{4-7\delta_2}{7} \cdot \frac{49\delta_2}{8-14\delta_2} - \frac{7\delta_2}{2} \\ &= \frac{7\delta_2}{2} - \frac{7\delta_2}{2} = 0, \end{aligned}$$

thus the left-hand side is $o(1)$. As a consequence, $\Delta(\mathcal{N}[v_1], \mathcal{N}[v_2]) = \Delta(N_1, N_2) \geq \Delta(G[X_1], G[X_2])(1 + o(1)) \geq \omega(1)$. Therefore, conditional on $A_1 \cap A_2 \cap A_3$, with probability $1 - o(n^{-2})$, v_1 and v_2 are locally asymmetric. \square

6. CONCLUSION

In this paper, we have introduced the concept of local symmetry in graphs, and studied the asymptotic presence and absence of this property in the Erdős-Rényi random graph model, rigorously establishing regimes for either behavior to emerge a.a.s.. It is important to note that, in this model, one can find locally symmetric graphs with high probability even in regimes of unrealistically high average degree, close to $n^{1/3}$. A natural question to ask is whether this behavior persists when looking at k -local symmetry for $k > 1$. To answer this question would require a deeper understanding of the combinatorial aspects of this problem.

One might also ask whether, and to which extent, real-world networks exhibit local symmetry. This would certainly depend on both the nature of the network and its formation process. However, for certain classes of networks, the extant literature allows us to develop some intuition about what answer to expect. For instance, in the social network literature, a number of recent works have attempted to explore the limits of network anonymization [13] [19]. In particular, a technique known as *percolation graph matching*, or PGM [23], has been successfully used to match common nodes in the Twitter and Flickr networks [19], thus allowing knowledge of one network to be used to break the anonymity of the other. Intuitively, this suggests that each node in these social networks can be uniquely identified structurally within them, from which we would conclude that the networks are globally asymmetric. However, the fact that the PGM technique works by exploiting local neighborhoods would also indicate that these networks exhibit some kind of local asymmetry. Further investigation of this matter would allow us to start exploring the applicability of the local symmetry concept to real-world networks.

It should be noted that, unlike the $G(n, p)$ random graphs, real-world networks exhibit much higher structural diversity of vertices. For instance, networks such as the Internet [10], the Web [5], and scientific collaboration networks [20] are believed to have heavy-tailed degree distributions. This diversity can lead to a similarly diverse behavior regarding global and local symmetry. It is known that most real-world networks are globally symmetric, but the automorphism group of these networks is due to a large number of small subgraphs which are themselves symmetric and comprise vertices with small degrees [18], with high degree vertices being asymmetric to any other vertex in the network. While a simple metric,

such as the fraction of vertices with at least one globally or locally symmetric counterpart, would begin to shed a light on this phenomenon, more fine-grained symmetry metrics are desired to capture it in more detail.

REFERENCES

1. Lars Backstrom, Cynthia Dwork, and Jon Kleinberg, *Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography*, Commun. ACM **54** (2011), no. 12, 133–141.
2. Norman Biggs, *Algebraic graph theory*, 2nd ed., Cambridge University Press, 1993.
3. Béla Bollobás, *Random graphs*, second ed., Cambridge University Press, Cambridge, UK, 2001.
4. Stephen P. Borgatti and Pacey C. Foster, *The network paradigm in organizational research: A review and typology*, Journal of Management **29** (2003), 991–1013.
5. Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener, *Graph structure in the web*, Comput. Netw. **33** (2000), no. 1-6, 309–320.
6. Ho-Lin Chen, Hsueh-I Lu, and Hsu-Chun Yen, *On maximum symmetric subgraphs*, Graph Drawing, Springer Berlin Heidelberg, 2001, pp. 372–383.
7. Elizabeth M. Daly and Mads Haahr, *Social network analysis for routing in disconnected delay-tolerant manets*, 8th ACM MobiHoc, 2007, pp. 32–40.
8. J. Elbert Simões, D. R. Figueiredo, and V. C. Barbosa, *Power-law decay of the degree-sequence probabilities of two random graphs with application to graph isomorphism*, ArXiv e-prints (2016).
9. Paul Erdős and Alfred Rényi, *Asymmetric graphs*, Acta Mathematica Academiae Scientiarum Hungarica **14** (1963), 295–315 (English).
10. Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos, *On power-law relationships of the internet topology*, SIGCOMM Comput. Commun. Rev. **29** (1999), no. 4, 251–262.
11. H. M. Finucan, *The mode of a multinomial distribution*, Biometrika **51** (1964), no. 3/4, pp. 513–517 (English).
12. Matthias Grossglauser and Patrick Thiran, *Networks out of control: Models and methods for random networks*, 2014.
13. Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis, *Resisting structural re-identification in anonymized social networks*, Proc. VLDB Endow. **1** (2008), no. 1, 102–114.
14. Derek A. Holton and John Sheehan, *The Petersen graph*, Cambridge University Press, 1993.
15. Jeong Han Kim, Benny Sudakov, and Van H. Vu, *On the asymmetry of random regular graphs and random graphs*, Random Structures and Algorithms **21** (2002), no. 3-4, 216–224.
16. Philip Klein and Neal E. Young, *On the number of iterations for dantzig-wolfe optimization and packing-covering approximation algorithms*, SIAM Journal on Computing **44** (2015), no. 4, 1154–1172, Journal version of [1999]. Published online Aug. 2015.
17. Jon M. Kleinberg, *Authoritative sources in a hyperlinked environment*, J. ACM **46** (1999), no. 5, 604–632.
18. Ben D. MacArthur, Rubén J. Sánchez-García, and James W. Anderson, *Symmetry in complex networks*, Discrete Applied Mathematics **156** (2008), no. 18, 3525 – 3531.
19. Arvind Narayanan and Vitaly Shmatikov, *De-anonymizing social networks*, Security and Privacy, 2009 30th IEEE Symposium on, IEEE, 2009, pp. 173–187.
20. M. E. J. Newman, *The structure of scientific collaboration networks*, Proceedings of the National Academy of Sciences of the United States of America **98** (2001), no. 2, 404–409.
21. Narciso Pizarro, *Structural identity and equivalence of individuals in social networks: Beyond duality*, International Sociology **22** (2007), no. 6, 767–792.
22. Olaf Sporns, *Networks of the brain*, 1st ed., The MIT Press, 2010.
23. Lyudmila Yartseva and Matthias Grossglauser, *On the performance of percolation graph matching*, Proceedings of the First ACM Conference on Online Social Networks (New York, NY, USA), COSN ’13, ACM, 2013, pp. 119–130.